Cloud computing, conferences and communities

Wouter Los discusses the importance of biodiversity research in modern sciences and its need for dissemination and global interaction supported by cloud services. He also explains how cloud computing is playing a central role in modern biodiversity research infrastructures

How has the design and implementation of research infrastructures to support biodiversity and ecosystems evolved?

As always, innovative technologies are opening up new areas of research and are assisting the testing of new hypotheses. Scientific developments in biodiversity research have often been limited by unsophisticated instrumentation or restricted localities. Despite this, much progress has been made in past centuries. For example, we have seen large undertakings by Carl Linneaus who sent his assistants out over the world, or marine expeditions to collect specimens of different species. This resulted in natural science collections acting as research infrastructures to serve visiting researchers with a condensed representation of the natural reality. We also see increasingly new facilities supporting research on our living environment.

In what ways has CReATIVE-B empowered scientific communities?

The supporting project partners originate from various communities in order to deliver input for the project's global workshops and digest the workshop results in technical and policy papers. We have partners representing large scientific associations of research institutes, companies with an interest in biodiversity, technical organisations from the public and private sector, and state and academic legal experts.

Could you comment of the success of the Global Biodiversity Informatics Conference (GBIC), which took place in Copenhagen, 2-4 July 2012?

Creative-B co-sponsored GBIC, as this conference brought together 100 invited experts from around the world to discuss how informatics can best meet the challenges posed by biodiversity science and policy. Some of the agreed conclusions include:

- Making best use of the huge potential for the public to become part of a global biodiversity knowledge network as both contributors and beneficiaries, by using the latest technologies, social networks and local/indigenous knowledge
- Capturing through all available technologies the complexity of interactions among species – for example predators/prey, parasites/hosts and pollinators – as well as their traits. The technologies will include acoustic monitoring and remote sensing, and will help analyse these interactions to establish their importance in providing ecological services to people
- Greatly improving the capability to provide predictive modelling across different scales, estimating the impact of specific environmental changes on biodiversity for any point on Earth, and the resulting disruption of ecological services to people and communities
- Expanding the current network of linked data from species names and museum collections up to satellite images of ecosystems and down to DNA in microorganisms
- Shining a light on hitherto hidden layers of biodiversity, for example using gene sequencing capabilities to understand the millions of kinds of microbes inhabiting the air, oceans, soils and higher organisms throughout the world, and their role in controlling the life support systems of the planet

Share and share-alike

Biological data helps explain the mysteries of the biodiversity and ecosystems that surround us. However, present fragmentation of data and analytical tools requires the cooperation of research e-infrastructures at the global scale – a situation that **CReATIVE-B** is helping to address

BIOLOGICAL DATABASES ARE libraries on the diversity and functions of life, from genomics up to whole ecosystems. This data is complex by its very definition: each observation or measurement is a single snapshot of a gene, an organism or habitat. It may explain a process or aspect of a system but, as biology is dynamic, huge volumes of linked data are necessary to accurately build a picture of changing systems and their interactions.

Given the huge efforts necessary for the production of quality data, scientists can benefit from sharing their data, integrating their efforts and using each other's work to create more reliable datasets, allowing the completion of better analyses and formation of firmer conclusions. In this context it is unsurprising that modern biological research has attempted to utilise new and evolving informatics techniques to facilitate sharing of data and the growing number of methods (algorithms and software tools) to support the analysis and modelling of the data.

It is this crucial process of effective data and tools sharing that the CReATIVE-B project, coordinated by Wouter Los, aims to achieve on a global scale with informatics and integration. There are three main challenges facing the development of e-infrastructures in biodiversity and ecosystem research: firstly, the production of new techniques and technologies which allow the more accurate and faster collection of data – ideally an automated process; secondly the provision of sufficiently large datasets to allow scientists to detect ecological or evolutionary patterns and processes not observable in smaller samples; and finally the provision of an infrastructure which allows scientists to control automated sensors, manage the data produced and analyse the results.

THE IMPORTANCE OF QUALITY

Despite a clear and unambiguous idea of what is required, achieving this is not a simple process: "The challenge is to operate a common and trusted infrastructure that can deliver the requirements with quality," points out Los. There is no benefit to large and sprawling datasets if scientists cannot rely on the quality of the measurements. This need for quality translates into a need for continuity and clarity in data collection, representation and integration. The problem can, in part, be solved through the introduction of widespread and well understood protocols and standards regarding the collection of relevant data. This need for widespread agreement dictates the global scale of the project.

It may be a cliché – but also a basic truth – that Nature does not adhere to state borders. Consequently, these proposed infrastructures must be international and cooperative. Cooperation is vital not only for the linking of infrastructure capabilities but also the division of tasks and the generation of research-relevant economies of scale. The consortium is studying the legal and technical issues that surround international cooperation: "The final target is to agree on a common global roadmap," Los indicates. The team then hopes this roadmap can facilitate and guide the cooperating research infrastructures and inform policy makers on how best to support such an endeavour.

What were some of the major highlights from this event?

The capabilities discussed by the participants at GBIC – who came from a range of disciplines including biodiversity science, policy and informatics – will result in an outlook document. It will set priorities for biodiversity informatics for the coming decade with a view to establishing an effective and agile system of forecast and rapid response – equivalent to weather forecasting or earthquake detection. A number of specific areas were identified for development in the outlook, each to include achievable outcomes over a five to 10-year timeline, building on and integrating many existing initiatives and contributing to the overall vision of a global biodiversity intelligence system.

How is cloud computing being implemented by CReATIVE-B to improve efficacy of research infrastructures that support biodiversity and ecosystems?

Biodiversity researchers increasingly want to tackle questions dealing with complexity of genetic, species, population and ecosystem diversity, as well as their mutual interactions and with the non-biological environment. Indeed the growing capabilities of research infrastructures can increasingly meet these demands. However, we cannot assume that researchers are also interested in the technical complexity behind the infrastructure services.

The solution is a virtual environment allowing scientists to intuitively navigate, find data, compose workflows, run computation and see visualised results. Such a user-friendly service should refrain from only pre-defined models from a few ecological schools with their preferred scientific paradigms. The first step in incorporating flexibility is by offering cloud-based services as a menu to choose from. The second is the development of a service through which interested users can compose their own menu by selecting and bringing together required elements from a repository of tools. With these approaches we can create user-driven development of a digital environment.

CREATIVE-B

COORDINATION OF RESEARCH E-INFRASTRUCTURES ACTIVITIES TOWARD AN INTERNATIONAL VIRTUAL ENVIRONMENT FOR BIODIVERSITY

OBJECTIVES

- Cooperation of the Global Scientific Communities dealing with the construction, operation and use of large infrastructures and facilities for Biodiversity and Ecosystems research
- Exchange of expertise on technological construction and operation technologies, with a focus on interoperability by mapping common reference models
- Development of a common view and coordination actions with respect to the governance and management aspects of large-scale distributed biodiversity research infrastructures (biodiversity RIs) and facilities
- Organisation of six workshops with the cooperating initiatives and contribution to the Global Biodiversity Informatics Conference

PARTNERS

Universiteit van Amsterdam, The Netherlands • **Cardiff University**, UK • **maatG**, Spain • **Consiglio Nazionale delle Ricerche**, Italy • **Universidad de Alcalá**, Spain • **Comunità Ambiente**, Italy • **Centre National de la Recherche Scientifique – Institut des Grilles et du Cloud**, France

FUNDING

EU Seventh Framework Programme (FP7) – contract no. 284441

CONTACT

Yannick Legre
Research Manager

CNRS
Institut des Grilles et du Cloud (IdGC)
c/o Laboratoire de Physique Corpusculaire
Campus des Cezeaux
24 avenue des Landais
BP 80026
63171 Aubiere CEDEX, France

T +33 473 405 373
E yannick.legre@idgrilles.fr

www.creative-b.eu

WOUTER LOS was Director of the Zoological Museum at the University of Amsterdam, resulting in his interest for applying large-scale biodiversity data in research. Currently he is coordinating the LifeWatch infrastructure for biodiversity and ecosystem research and a few projects dealing with the interoperability of environmental research infrastructures.

IMPRESSIVE INPUT

The project is starting to gather pace and a number of major research and data organisations have signed up including LifeWatch, Atlas of Living Australia, Chinese Academy of Sciences, Reference Centre on Environmental Information (Brazil), DataOne (USA, see p15), Global Biodiversity Information Facility (Denmark), GEO Biodiversity Observation Network and the South African National Biodiversity Institute. The common hope is that their e-infrastructure cooperation can better serve scientists with accelerated understanding of biological processes, a greater output of meta-analysis and the identification of holistic patterns only observable in large-scale datasets.

> Data is powerful, but its power increases exponentially with the development of novel tools to understand them

Despite the improvements that such cooperation will provide, the actual collection of data outside in nature is still time-consuming and often challenging. For these reasons, the team aims to empower the relatively underused resource of citizen scientists. An infrastructure of humans that allows data collection by public observation could lead to a dramatic increase the amount and diversity of collected data. New technologies should be developed to support their data collection, for example with interactive mobile devices. These contributors could be vital in future interactive datasets and also give members of the public a very tangible sense of involvement and ownership of scientific development and dissemination.

POINT OF USE

Ensuring the network meets user needs is a major focus for the group. To this end, the researchers have to consider the range of uses that exist: "The evolving requirements of user communities are taken into account, with special attention to national and regional differences," Los highlights. The project itself has a lifespan of three years; in order to ensure that continuity and cooperation endures beyond this point it has been vital to set up a 'High Level Stake Holders Board' comprising selected members of each research infrastructure. These board members will further promote the measurable benefit of their cooperation after the current funding period.

While the development of the infrastructure designed by CReATIVE-B requires common priorities and shared informatics and computing expertise, the success of the idea depends on the ability to 'sell' it. As such one of the efforts of the team is the dissemination and communication of their concepts and goals, organising workshops and conferences for interested researchers. Los is clear about the importance of these workshops: "The workshops around the world are expected to contribute to drafting a common roadmap on community priorities, a strategy toward enhanced interoperability, and on the related legal and governance implications". The workshops will provide Los and his colleagues with an opportunity to understand geographical differences in scientific and cultural opinion which impact on both the input and end point of their infrastructure services. They see their work as having a crucial role to play in facilitating the bridging of these cultural and scientific differences.

Data is powerful, but its power increases exponentially with the development of novel tools to understand them. Consequently, scientists have an interest in contributing to research e-infrastructures and the pooling of services. By enhancing access to diverse sets of data and providing technical services allowing users to select and integrate selected data in their preferred processing tools, the CReATIVE-B project is proving itself to be of global importance. Without these coordination efforts, such efficient access to biological data would remain impossible.